MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1665–1680 doi: 10.1111/mec.13563

EPIGENETIC STUDIES IN ECOLOGY AND EVOLUTION

Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients

PAUL F. GUGGER,*[1] SOREL FITZ-GIBBON,*† MATTEO PELLEGRINI†‡ and VICTORIA L. SORK*§

*Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, USA, †Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095, USA, ‡Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA, §Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-1496, USA

Abstract

DNA methylation in plants affects transposon silencing, transcriptional regulation and thus phenotypic variation. One unanswered question is whether DNA methylation could be involved in local adaptation of plant populations to their environments. If methylation alters phenotypes to improve plant response to the environment, then methylation sites or the genes that affect them could be a target of natural selection. Using reduced-representation bisulphite sequencing (RRBS) data, we assessed whether climate is associated with variation in DNA methylation levels among 58 naturally occurring, and species-wide samples of valley oak (*Quercus lobata*) collected across climate gradients. We identified the genomic context of these variants referencing a new draft valley oak genome sequence. Methylation data were obtained for 341 107 cytosines, of which we deemed 57 488 as single-methylation variants (SMVs), found in the CG, CHG and CHH sequence contexts. Environmental association analyses revealed 43 specific SMVs that are significantly associated with any of four climate variables, the majority of which are associated with mean maximum temperature. The 43 climate-associated SMVs tend to occur in or near genes, several of which have known involvement in plant response to environment. Multivariate analyses show that climate and spatial variables explain more overall variance in CG-SMVs among individuals than in SNPs, CHG-SMVs or CHH-SMVs. Together, these results from natural oak populations provide initial evidence for a role of CG methylation in locally adaptive evolution or plasticity in plant response.

Keywords: DNA methylation, environmental association analysis, epigenomics, local adaptation, *Quercus lobata*, reduced-representation bisulphite sequencing

Received 8 October 2015; revision received 25 January 2016; accepted 27 January 2016

Introduction

DNA (cytosine) methylation may play an important role in transposon silencing, transcriptional regulation and cell differentiation and consequently shapes the phenotypes of organisms (Zilberman *et al.* 2007; Cokus *et al.* 2008; Law & Jacobsen 2010; Duncan *et al.* 2014; Eichten *et al.* 2014; Schmitz 2014; Widman *et al.* 2014). Foundational work in *Arabidopsis thaliana* and other plants demonstrates that levels of DNA methylation at sites throughout the genome vary substantially among individuals and populations (Herrera & Bazaga 2010; Schmitz *et al.* 2013). Active debate surrounds the question of whether (and to what extent) this epigenetic variation could be involved in local adaptation, either as a mechanism acting independently of genetic variation or as an intermediary through which genetic varia-

Correspondence: Victoria L. Sork, Fax: +1 310 206 0484; E-mail: vlsork@ucla.edu

[1]Present address: Appalachian Laboratory, University of Maryland Center for Environmental Science, 301 Braddock Road, Frostburg, MD 21532, USA

tion can affect phenotype (Grossniklaus *et al.* 2013; Cortijo *et al.* 2014; Dubin *et al.* 2015). Furthermore, some methylation polymorphism can be induced under environmental stress and may or may not be inherited in subsequent generations, offering potential mechanisms for phenotypic plasticity or Lamarckian evolution (Jablonka & Raz 2009; Verhoeven *et al.* 2010; Verhoeven & van Gurp 2012; Duncan *et al.* 2014). Moreover, in plants, many DNA methylation patterns appear to be heritable, but the nature and stability of the heritability across generations is not well established (Schmitz *et al.* 2011; Becker & Weigel 2012; Chodavarapu *et al.* 2012; Grossniklaus *et al.* 2013; Zhang *et al.* 2013; Wilschut *et al.* 2015).

The potential role of DNA methylation as an epigenetic mechanism involved in either local adaptation or plasticity is an area of growing interest in natural contexts because of its implications for understanding short-term responses to environmental change as well as long-term evolution (Bossdorf *et al.* 2008; Franks & Hoffmann 2012; Liu 2013; van der Graaf *et al.* 2015). Because DNA methylation state has a substantially higher mutation rate than DNA sequences (Jablonka & Raz 2009; Becker *et al.* 2011; Schmitz *et al.* 2011), it may provide a particularly useful mechanism for generating phenotypic variation that would allow a more rapid response to environmental change than would genetic mechanisms (Liu 2013). Multiple studies suggest that methylation can be induced by the local environment (Verhoeven *et al.* 2010; Uthup *et al.* 2011; Herrera & Bazaga 2013; Rico *et al.* 2014; Yakovlev *et al.* 2014). In principle, locally adapted phenotypes emerge more quickly and populations adapt more rapidly when both epigenetic and genetic variations are acted upon by natural selection compared to when only genetic variation is involved (Klironomos *et al.* 2013; Kronholm & Collins 2015).

Studies of methylation variation in natural populations and its relation to environmental factors can provide valuable information on these and other topics. Several ecological studies have found associations between variation in DNA methylation with herbivory (Herrera & Bazaga 2011; Holeski *et al.* 2012), climate stress (Verhoeven *et al.* 2010; Dowen *et al.* 2012; Rico *et al.* 2014; Nicotra *et al.* 2015) as well as other environmental factors that differ among populations (Foust *et al.* 2013, 2016). Furthermore, evidence from Norway spruce suggests that epigenetic regulation affects phenotypes that can be important in local adaptation, such as flowering time and climate adaptation (Yakovlev *et al.* 2011, 2012). Because many of these studies generated their evidence using methylation-sensitive AFLP (MSAP) loci or other methods that do not provide the genomic sequence context of the methylation patterns, they do not allow the association of methylation variants with specific genes and thus do not provide the potential mechanism of their involvement in adaptation. As it is now feasible to create a draft genome sequence for nonmodel species, ecological studies of the association between epigenetics and environment will benefit from a genomic sequencing approach that identifies the locations of methylation sites and the genes that surround them.

A recent whole-epigenome bisulphite sequencing study of local adaptation in natural Swedish accessions of *Arabidopsis* revealed the potential power of understanding DNA methylation in its genomic context (Dubin *et al.* 2015). This study showed that methylation in the CHH sequence context (where H is A, C or T) is significantly associated with greenhouse temperature and that these variants are primarily found in transposons and related to transposon silencing. This variation was shown to have an underlying genetic basis, especially due to a single previously known gene in *trans* (Shen *et al.* 2014). Conversely, they reported that methylation in the CG context is largely found in genes, variable among accessions, correlated strongly with latitude and climate variables, positively correlated with expression of those genes and, in some cases, genetically based. These results and others suggest that gene body CG methylation variation could be associated with local adaptation by natural selection, either directly or through underlying genetic variation, even though its relationship with phenotype is not well understood (Schmitz *et al.* 2013; Dubin *et al.* 2015; Wang *et al.* 2015). In contrast, CHG methylation was not shown to be sensitive to growing conditions and exhibited only modest correlations with environmental variables. These findings are consistent with known differences in the biochemical pathways that maintain methylation in these three different contexts in plants (Law & Jacobsen 2010), as well as previously reported roles of different DNA methylation types in transposons versus genes (Zilberman *et al.* 2007; Cokus *et al.* 2008; Takuno & Gaut 2012).

A detailed investigation of the role of DNA methylation in response to climate variation is of particular interest in long-lived trees because their long generation time limits the ability of populations to respond to rapid environmental changes through genetic mechanisms (Franks & Hoffmann 2012; Bräutigam *et al.* 2013). To explore whether DNA methylation might be a mechanism of tree response to the environment, we initiated a landscape epigenomics study of methylation in valley oak (*Quercus lobata*), a long-lived California endemic tree species with a genome size of ~800 Mb (Plomion *et al.* 2016; V.L. Sork, S. Fitz-Gibbon, M.W. Crepau, P.F. Gugger, C.H. Langley, M. Pellegrini, D. Puiu & S.L.

Salzberg, unpublished data). Our previous work showed that CG single-methylation variants (CG-SMVs, also referred to as single-methylation polymorphisms, SMPs) are unusually differentiated among three populations, more so than the average SNP or CHG-SMV (Platt *et al.* 2015). As a result, we concluded that CG-SMVs are more likely to be associated with local adaptation than CHG methylation variants, as reported for *Arabidopsis* (Schmitz *et al.* 2013; Dubin *et al.* 2015). The findings of that previous study provide compelling evidence based on population differentiation, but they do not address which environmental selective pressure is driving these differences. If climate were the driving force, we would expect that some methylation variants should be strongly associated with climate gradients, after factoring out the population or kinship structure. Furthermore, if CG-SMVs are generally more important in climate adaptation, we hypothesize that CG-SMVs will show stronger overall associations with climate gradients on the landscape than other types of SMVs and SNPs.

In this study, we assess the extent to which DNA methylation is associated with climate in valley oak using reduced-representation bisulphite sequencing (RRBS) informed by a valley oak draft reference genome sequence, Version 0.5 (V.L. Sork, S. Fitz-Gibbon, M.W. Crepau, P.F. Gugger, C.H. Langley, M. Pellegrini, D. Puiu & S.L. Salzberg, unpublished data) and an annotated reference transcriptome (Cokus *et al.* 2015). By sampling individuals across the species range that includes diverse climate environments, we assess the opportunity for local adaptation in natural populations through four specific objectives. First, we describe the patterns of methylation composition and distribution across the genome, focusing on different contexts of DNA methylation. Second, we conduct an environmental association (outlier) analysis of CG, CHG and CHH methylation levels with climate gradients to test the prediction that climate is a driving force in population differentiation and local adaptation. Third, we assess the general importance of CG methylation and identify which climate variables are most associated with methylation variation through multivariate association analyses of each CG, CHG and CHH methylation levels and SNP allele frequencies with climate variables. Fourth, we explore the genomic context of climate-associated SMVs to determine the position of these SMVs relative to genes whose expression may be associated with methylation. Despite the potential benefits to rapid response to environmental change offered by methylation, this study is among the first to test for evidence of epigenetically based adaptation to environment and determine potential genes that are mediating the phenotype in natural populations of a long-lived tree.

Methods

Study system

Valley oak, *Quercus lobata* Née (Fagaceae), is a widespread endemic oak of California occurring along the foothills of the Coastal and Sierra Nevada ranges (Fig. 1). In contrast to oak species in eastern North America and Europe that went through recent postglacial expansion, valley oak's distribution lies in a nonglaciated region of California and has likely remained stable through recent glacial cycles, leading to high local genetic diversity, associations of genetic variation with climate variables, and a strong potential for local adaptation to environment (Grivet *et al.* 2006; Sork *et al.* 2010, 2016; Gugger *et al.* 2013). Our previous work on epigenetic variation across three populations provides evidence that CG methylation could be either a marker for loci involved in local adaptation or directly involved in local adaptation (Platt *et al.* 2015).

Sampling

During October 2012, we sampled mature leaves from *Quercus lobata* at each of 58 localities spread throughout its entire distribution, ensuring sampling of the entire climate gradient (Fig. 1; Table S1, Supporting information). Because we wanted to maximize the range of climate environments in our study, we sampled only one individual from each site, which is sufficient for the regression-type analyses of this study.

Library preparation

We extracted total genomic DNA from the leaves using a prewash protocol (Li *et al.* 2007; Gaddis *et al.* 2014) followed by Qiagen DNeasy Plant extraction kits. Reduced-representation bisulphite sequencing (RRBS) libraries were prepared following methods established in the Pellegrini Lab, which are modifications of the protocol of Feng *et al.* (2011). Briefly, total genomic DNA was digested with *Msp*I (CCGG) and then end-repaired and adenylated with Klenow fragment (3′ to 5′ exo−). Unique Illumina TruSeq adapters were ligated to fragments in each library. We targeted fragments of 200–500 bp (including ~120 bp of adapter sequence) using AMPureXP bead-based size selection. Libraries were treated with sodium bisulphite (EpiTect, Qiagen) to convert unmethylated cytosines to uracil, which are read as thymine during sequencing. The resulting 58 reduced-representation libraries were amplified by PCR with Illumina primers, pooled in batches of 12 and single-end, 100 bp sequenced in 5 lanes within the same

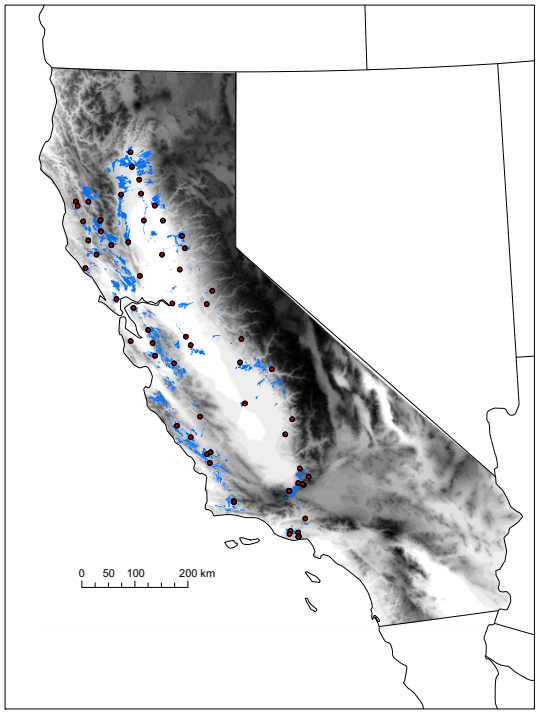

Fig. 1 Map of sample sites (red points) in relation to the distribution of valley oak (blue) and topography (grey scale).

flow cell of an Illumina HiSeq 2000 v3 according to the standard Illumina protocol.

SMV and SNP calling

Illumina reads were filtered to remove those failing the Illumina chastity test, converted from QSEQ to FASTQ format applying EAMSS quality score correction and demultiplexed by sample. Each sample was then analysed with BS-Seeker2 (Guo *et al.* 2013) to measure cytosine methylation in CG, CHG and CHH contexts. Specifically, we ran BS-Seeker2 with the Bowtie 2-2.2.5 aligner against a draft valley oak reference genome (ValleyOak0.5) that has haplotypes collapsed (S. Fitz-Gibbon, D. Puiu, P.F. Gugger & V.L. Sork, unpublished data; Sork *et al.* in prep.), which was indexed with a fragment size range of 100–480 bp (sequences are publicly available through NCBI PRJNA308314 and through http://valleyoak.ucla.edu). Other settings were left at their defaults, except that reads were aligned end to end after trimming adapter sequence, up to five mismatches were allowed, and methylation calls were only retained for cytosines with at least 10× coverage.

The resulting cytosines with methylation calls were filtered to include only those with less than 10% missing data across the 58 trees and that were considered variable across samples. Unlike SNPs, which are given by discrete nucleotide calls, methylation levels are measured as continuous values that capture the fraction of methylation at a site (representing the fraction of cells containing methylation at a locus). We defined single-methylation variants (SMVs) as those with a 10% minimum range of variation in per cent methylation across samples. We additionally truncated individual per cent methylation calls that were deemed outliers. Specifically, we defined a range based on the 10th and 90th percentiles and considered outliers those that were 1.5 times more extreme than those values. These outliers were converted to the cut-off values themselves, resulting in 'winsorized' data. Linear models are sensitive to outliers (i.e. potentially high-leverage points), and thus, we did this to minimize the false-positive rate of associations.

Because RRBS generates sequences, we are able to also call SNPs for contexts that are not confounded with potential cytosine methylation. Target genomic intervals for variant calling were identified using `CallableLoci` in GATK 3.3 (McKenna *et al.* 2010). To avoid spurious calls at the edges of the RRBS target regions, the target intervals were trimmed by three bases on each end. bwa-meth 0.10 (Pederson *et al.* 2014) was used to align the reads for input to Bis-SNP (Liu *et al.* 2012). Bis-SNP's `BisulfiteGenotyper` was run with the following parameters: `-stand_call_conf 4 -stand_emit_conf 0 –maximum_read_cov 100000 –mm40 10 –trim5 2 –trim3 2 –mmq 10 –mbq 15 –useBAQ`. GATK's `VariantFiltration` and `SelectVariants` were used to remove calls for sites with low depth of coverage (DP < 4), restrict to diallelic sites and filter variants based on the following GATK variant annotation cut-offs, QD < 2.0, MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0 and AF < 0.1.

Genomic context of SMVs and SNPs

We identified the genomic context for each variant based on our draft genome sequence (S. Fitz-Gibbon, D. Puiu, P.F. Gugger & V.L. Sork, unpublished data) and our published transcriptome assembly (Cokus *et al.* 2015). As we do not yet have a carefully annotated genome sequence, we used BLASTx against the NCBI nonredundant database to identify genes in and around regions with climate-associated SMVs. Only matches with compelling evidence of homology were kept, as indicated in the expect score and per cent matched.

To identify genomewide patterns of SMV contexts, we used our transcriptome assembly to produce a conservative set of gene annotations on our draft genome, which allowed us to compare frequencies of genic versus intergenic SMVs and estimate frequencies around transcription start sites. Specifically, from the 83 644 transcriptome contigs in our published transcriptome assembly, we identified 13 159 high-quality contigs by

requiring that the annotation included UTRs, start and stop codons and no introns. These 13 159 transcriptome contigs were mapped to the 40 158 contigs from our draft genome assembly with `Gmap --batch 5` (Wu & Watanabe 2005). Alignments for the 12 599 transcriptome contigs with matches were refined by passing pairs of transcriptome contigs and their matching genomic regions, plus and minus 50 kb, to SIM4DB (Walenz & Florea 2011) with default parameters. Of the resulting alignments, 7956 transcriptome contigs were determined to have full-length matches to the genomic contigs, which were the only annotations kept for subsequent use. Thus, our transcriptome-based annotation of the genome covers only a fraction of the true number of genes; however, it is a high-confidence subset. Locations of methylation sites with respect to annotated gene models were determined using the `count` function of SNPEFF 4.1k (Cingolani *et al.* 2012). Patterns of methylation and density of Cs and RRBS fragments around transcription start sites were determined using a slightly smaller subset of 5799 gene annotations for which the direction of coding (based on observed splice signals) had been determined by SIM4DB. BEDTOOLS 2.19.1 (Quinlan & Hall 2010) was used to facilitate creation of sliding windows around the transcription start sites as well as for counting intersecting data.

Climate association analyses

To identify specific SMVs that have significant associations with the climate, which could be indicative of natural selection or a strong influence of climate on natural variation at that locus (Yoder *et al.* 2014; Gugger *et al.* in press), we used linear mixed models implemented in PYLMM (http://genetics.cs.ucla.edu/pylmm/index.html). PYLMM implements the same model as popular software for linear mixed models, such as EMMAX (Kang *et al.* 2010), but the input formats are not restricted to those of SNP data. These tests adjust for pairwise relatedness, which is the correlation structure among samples, using a kinship matrix estimated from the data (Kang *et al.* 2008; Sul & Eskin 2013) and then analyse individual SMVs for significant associations with climate variables. We used the methylation data to estimate the kinship matrix because we determined that it captures the relatedness of individuals and controls the rate of false positives better than using the SNP data for estimating kinship. As evidence, we present QQ plots demonstrating an inflated number of significant tests when using SNPs rather than methylation to estimate kinship (Fig. S1, Supporting information). Furthermore, all of the SMVs shown to be significant using SMV-based kinship are also the most highly significant when using SNP-based kinship (Fig. S2, Supporting information). Similar observations have been reported in other contexts and are attributed to underlying association of methylation variation with genetic variation, as well as the better statistical properties of using the same data type for the correlation matrix and dependent variables (Orozco *et al.* 2015). Loci with >10% missing data were excluded from the PYLMM analyses, and only sufficiently variable loci (SMVs), as defined above, were included in the analysis to minimize false positives. Furthermore, we adjusted *P*-values to *Q*-values using the false discovery rate method (Storey & Tibshirani 2003) to account for multiple testing, as implemented in the R package QVALUE 2. For these analyses, we considered four climate variables with generally low-to-moderate correlations among each other that are thought to be important in shaping valley oak's geographic distribution and patterns of genetic variation (Sork *et al.* 2010; McLaughlin & Zavaleta 2012; Gugger *et al.* 2013): climatic water deficit (CWD; an integrated measure of water availability or stress considering rainfall, evapotranspiration and basin hydrology), mean minimum temperature of the coldest month (T_{min}) and mean maximum temperature of the warmest month (T_{max}) from 1950 to 1980 averages from the 270-m resolution California Basin Characterization Model (Flint *et al.* 2013), as well as growing season growing degree-days above 5°C (GSDD5) from a spline model of climate developed by the U.S. Forest Service (Rehfeldt 2006) as a measure of energy input available for tree growth.

In a parallel analysis to assess the overall association of different classes of SMVs and SNPs with climate on the landscape, we used a multivariate method called redundancy analysis (RDA), as implemented in VEGAN 2.3 (Oksanen *et al.* 2015) in R 3.1.2 (R Core Development). RDA is a constrained ordination method analogous to linear regression for cases that have multiple-dependent variables (SMVs) and multiple-independent variables (climate and spatial variables). We used it to test and quantify the overall contribution of climate and spatial variables to different types of cytosine methylation, including only sites without any missing data. The analysis was performed separately for each SMV context and for SNPs. Permutation tests with 999 permutations were used to assess the significance of the associations. Spatial variables included longitude (x), latitude (y) and elevation (z).

Results

DNA methylation and its sequence context

We obtained methylation data for 1.47 M cytosines spread across approximately 36 k RRBS fragments, of which 341 107 cytosines had less than 10% missing

data, and were mostly in the CHH context, followed by CG and CHG (Table 1). Methylation data are enriched near transcription start sites, in part because RRBS preferentially samples these regions (Figs S3 and S4, Supporting information). As is commonly seen in eukaryotic genomes, there is a spike in the number of CG sites around the transcription start sites and a dip in the average methylation level of these sites (Fig. S3, Supporting information). The spike in C density is also seen for CHG and CHH sites; however, the dip in average methylation is much less pronounced for those contexts. 12.5% of CG, CHG and CHH sites occur in genic regions, including exons, introns, 1 kb upstream and 1 kb downstream (Fig. S5, Supporting information). Methylation levels in all contexts are strongly skewed towards zero, but CG methylation is often found close to 100% (Fig. 2).

After selecting only SMVs with less than 10% missing data and defined by a 10% minimum range of variation, 26 286 CG-SMVs, 9758 CHG-SMVs and 21 404 CHH-SMVs remained for use in subsequent analyses (Table 1). We observed 18% of CG-SMVs, 10% of CHG-SMVs and 6% of CHH-SMVs fall within genes, suggesting genic enrichment of CG-SMVs compared to other SMV classes and all methylation sites (Fig. S5, Supporting inforamtion). These numbers are low due to our conservative mapping with only the most confident gene models. CG-SMV sites exhibit a bimodal distribution of methylation levels among loci, with peaks towards 0 and 100% methylation (Fig. 2). CHG-SMV sites are far more skewed towards zero methylation but have some sites with 100% methylation. CHH-SMVs have mostly low levels of methylation. There is little variation in these distributions among samples, and truncating the values of outliers had no discernible effect on these distributions (not shown). In addition, we identified 1810 total high-confidence SNPs with less than 10% missing data (1394 with complete data).

Table 1 Number and per cent of sites with methylation data by sequence context and the subset of those that are at least 10% variable among individuals (i.e. SMVs)

Methylation context	Sites (<10% missing data)	Per cent of total	SMVs	Per cent of total
CG	65 203	19	26 286	46
CHG	50 658	15	9758	17
CHH	225 246	66	21 404	37
Total	341 107		57 448	

Climate associations

Linear mixed models revealed 43 SMVs significantly associated with climate variables after accounting for kinship structure and multiple testing ($Q < 0.1$) (Table 2). A total of 38 of the 43 are significant associations between CG-SMVs and T_{max} (Fig. 3) and come from 19 RRBS fragments (Tables 2 and 3). Just three fragments account for nearly half of these CG-SMVs, which we highlight in regression plots and maps with T_{max} (Fig. 4). In one case, 9 CG-SMVs from a single 100-bp fragment (scaffold20751.23308, which is named in the format, contig.position) are significantly associated with T_{max} (Figs. 4 and 5). In all but one fragment where a CG-SMV is significant, other SMVs (especially CHG and CHH) in that fragment are not significant. One fragment has a CG-SMV, CHG-SMV and CHH-SMV significantly associated with T_{max} (Fig. 4). Methylation levels at sites within the same RRBS fragment have low-to-moderate correlations, with mean of 0.19 and range of 0.07–0.36 across contexts (Table S2, Supporting information; e.g. Fig. 5).

RDAs show that CG-SMVs and CHG-SMVs are significantly associated with climate and spatial variables ($P < 0.001$) (Table 4; Fig. S6, Supporting information). In contrast, CHH-SMVs are not significantly associated with climate and spatial variables ($P = 0.23$), even when the outlying individual (SW786) whose CHH methylation is unusually low is removed from the analysis ($P = 0.07$). The amount of variance explained by climate and spatial variables is higher for CG-SMVs (14.4%)

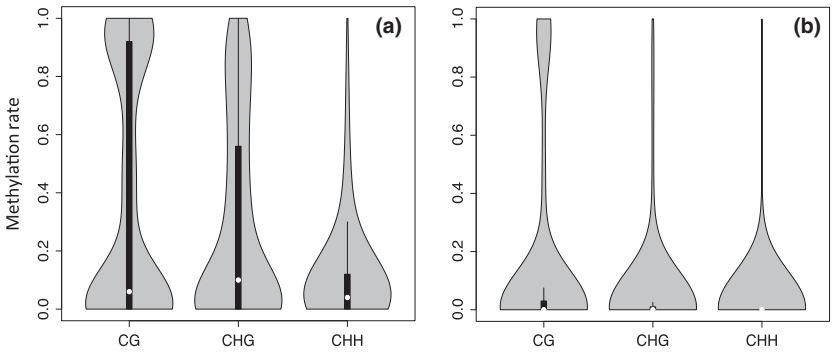

Fig. 2 Violin plots showing patterns of methylation levels of CG, CHG and CHH sites for (a) all cytosines with less than 10% missing data and (b) the subset of those with at least 10% variance among samples (i.e. SMVs). White dots represent median methylation level, and thick black bars represent the interquartile range. Mean methylation levels (not shown) reflect the greater number of highly methylated sites in CG-SMVs: CG = 0.38, CHG = 0.29, CHH = 0.11.

than for SNPs (13.3%) or for CHG-SMVs (13.6%) and CHH-SMVs (12.8%, not significant; Table 4).

Consistent with our single-site analyses, the RDA shows that T_{max} is most strongly associated with the first RDA axis; however, GSDD5 and CWD also have strong associations with one or both of the first two axes (Table 5). In contrast, the SNP RDA axes are more strongly associated with spatial variables and GSDD5 ($P < 0.001$).

Table 2 Summary of number of SMVs (number of fragments in parentheses) with significant environmental associations ($Q < 0.1$) for each climate variable investigated with linear mixed models

	Number of significant tests by climate variable			
Methylation context	CWD[1]	T_{min}[2]	T_{max}[3]	GSDD5[4]
CG		1	38 (19)	
CHG			1	1
CHH			1	1

[1]CWD = climatic water deficit.
[2]T_{min} = mean minimum temperature of the coldest month.
[3]T_{max} = mean maximum temperature of the warmest month.
[4]GSDD5 = growing season degree-days above 5°C.

Genomic context of climate-associated SMVs and nearby genes

The 43 climate-associated SMVs are enriched in or near genes (49%), even beyond the underlying enrichment of CG-SMVs in genic regions (24%; Fig. S5, Supporting information; hypergeometric test: $P = 0.0002$). As inferred from our draft genome sequence, 15 of 43 climate-associated SMVs are found within coding sequences (Table 3). Another 7 are within 1 kb of coding regions, and only 5 are more than 2.5 kb from detected coding regions. All of the climate-associated SMVs within 2.5 kb of coding regions are downstream.

The genes found to be near the SMVs with significant climate association include several thought to have a role in environmental stress response or adaptation (Table 3). For example, scaffold20751.23308, which contains nine methylated sites, is found in a coding sequence whose protein product has 60% identity to dehydration-responsive element-binding protein 1A found in the woody plants *Jatropha* and *Populus* with functional annotation from *Oryza sativa* and *Arabidopsis* (Okamuro *et al.* 1997; Dubouzet *et al.* 2003).

Discussion

Methylation patterns in species-wide valley oak samples provide evidence of response to climate and potential involvement in local adaptation. Our evidence is

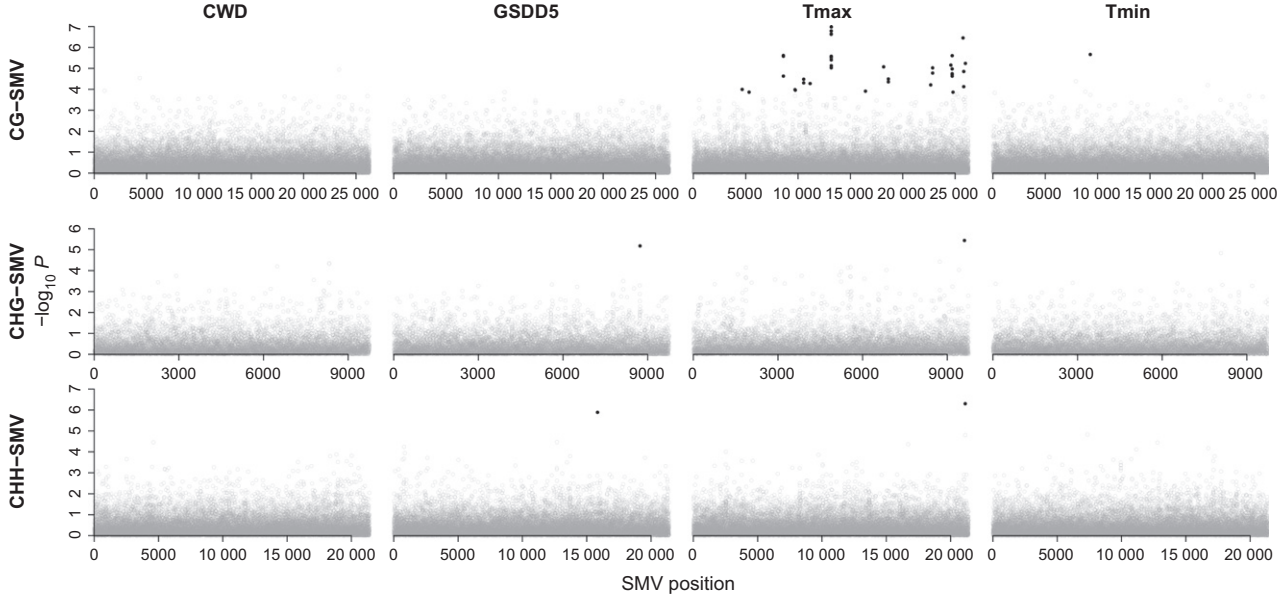

Fig. 3 Manhattan plots showing the −log*P* of a correlation between each single-methylation variant (SMV) (rows) and climate variables (columns): CWD, climatic water deficit; GSDD5, growing season degree-days above 5°C; T_{max}, mean maximum temperature of the warmest month; and T_{min}, mean minimum temperature of the coldest month. Black-filled points are statistically significant ($Q < 0.1$) after adjusting for multiple testing using the false discovery rate method. SMVs are arranged along the *x*-axes by contig (largest to smallest) and then by position within contig, following the draft genome sequence ordering.

Table 3 Genomic context of fragments containing SMVs with significant association to climate, including number of SMVs within fragment, closest gene, distance to closest gene (− means upstream) and putative protein product, sorted by significance level. The fragment label indicates the genomic contig and the position of the most highly associated SMV (contig.position). For fragments with multiple significant SMVs, the lowest *P*-values are listed. Functions indicated by the title of the BLASTx match are themselves often inferred from a homolog match and are speculative. Distances are in nucleotides and are measured from the closest end of the amino acid alignment

RRBS fragment	Climate association	Context	Climate association *P*-value (min)	Closest gene via BLASTx match	Distance from BLASTx-identified gene (bp)	BLASTx Expect	Per cent of nr database gene matched (amino acids)
scaffold20751.23308	T_{max}	CG × 9	1.1E-07	dehydration-responsive element-binding protein 1A [*Jatropha curcas*]	genic	2.E-60	87 (30–225 of 225)
C2279863.599	T_{max}	CG	3.5E-07	DUF247 [*Medicago truncatula*]	genic	1.E-45	76 (99–449 of 458)
C2243561.395	T_{max}	CG × 2; CHG; CHH	5.0E-07	none on 1.0 kb contig			
scaffold10809.8828	GSDD5	CHH	1.3E-06	uncharacterized protein LOC103697430 [*Phoenix dactylifera*]	2460	7.E-35	76 (39–218 of 236)
scaffold9024.922	T_{min}	CG	2.2E-06	hypothetical protein VITISV_004376 [*Vitis vinifera*]	10 513	0.E+00	47 (1061–2036 of 2080)
C2687893.1977	T_{max}	CG × 6	2.5E-06	mitochondrial outer membrane porin 4-like [*Solanum tuberosum*]	394	1.E-43	62 (105–276 of 276)
C2135867.734	T_{max}	CG	5.8E-06	heparan-α-glucosaminide N-acetyltransferase isoform ×1 [*Vitis vinifera*]	209	9.E-18	18 (366–446 of 455)
scaffold36086.1055	GSDD5	CHG	6.6E-06	β-Fructofuranosidase, soluble isoenzyme I-like [*Fragaria vesca* subsp. *vesca*]	860	6.E-146	79 (135–643 of 644)
C2716265.837	T_{max}	CG	7.0E-06	zinc finger protein, putative [*Ricinus communis*]	6	9.E-33	47 (258–490 of 493)
scaffold1594.15565	T_{max}	CG	8.4E-06	none on 35 kb contig			
scaffold27436.2760	T_{max}	CG × 2	9.4E-06	none on 13 kb contig			
scaffold11289.53220	T_{max}	CG × 3	2.3E-05	gag-pol precursor [*Castanea mollissima*]	−5440	2.E-118	49 (538–1082 of 1106)
scaffold7563.31879	T_{max}	CG × 2	3.2E-05	hypothetical protein Csa_4G076240 [*Cucumis sativus*]	genic	2.E-64	83 (51–297 of 297)
scaffold2576.38890	T_{max}	CG × 2	3.3E-05	short-chain dehydrogenase/reductase family 42E member 1-like isoform ×2 [*Citrus sinensis*]	1315	2.E-71	98 (1–468 of 478)
scaffold4607.51780	T_{max}	CG	5.3E-05	dehydRP-like protein [*Eucalyptus cladocalyx*]	genic	1.E-54	99 (2–217 of 217)
scaffold20147.3286	T_{max}	CG	6.1E-05	hypothetical protein JCGZ_15830 [*Jatropha curcas*]	−10 585	2.E-73	84 (27–177 of 179, on contig edge)
scaffold1450.104386	T_{max}	CG	1.0E-04	sodium/pyruvate cotransporter BASS2, chloroplastic [*Jatropha curcas*]	711	8.E-34	46 (183–372 of 410)
scaffold15759.63429	T_{max}	CG	1.0E-04	serine carboxypeptidase-like 45 [*Vitis vinifera*]	genic	2.E-33	98 (4–456 of 460)
scaffold12904.46759	T_{max}	CG	1.1E-04	GDSL-like lipase/acylhydrolase [*Medicago truncatula*]	genic	4.E-67	96 (13–370 of 370)
scaffold877.15134	T_{max}	CG	1.2E-04	pentatricopeptide repeat-containing protein At1g20230 [*Prunus mume*]	1460	0.E+00	99 (7–756 of 756)
scaffold13292.62729	T_{max}	CG	1.4E-04	polygalacturonase-inhibiting protein [*Pyrus communis*]	566	7.E-133	98 (7–330 of 330)
C2660447.1010	T_{max}	CG	1.4E-04	SKP1-like protein 1B [*Nicotiana tomentosiformis*]	305	3.E-22	85 (11–75 of 75)

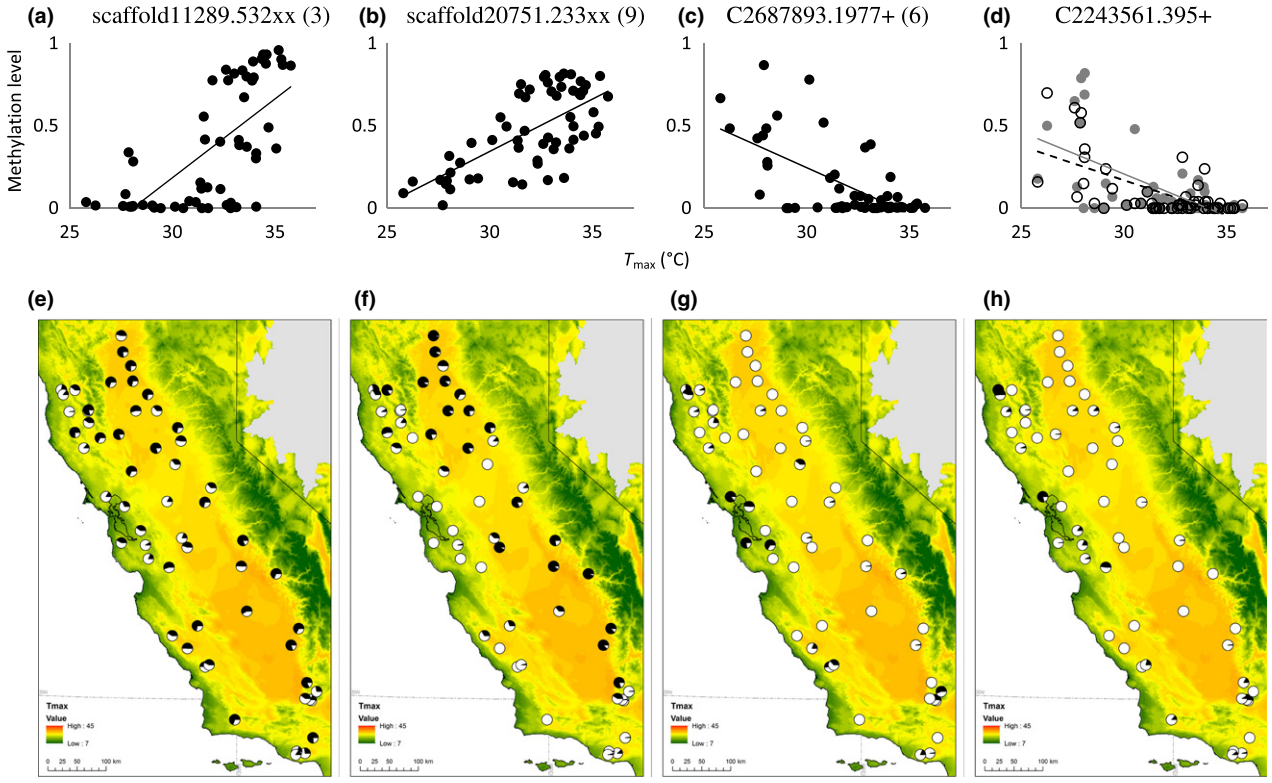

Fig. 4 Four examples of RRBS fragments with significant SMV–climate associations from mixed model regression in PyLMM. (a–d) Regression plots of mean level of methylation across methylated sites within RRBS fragments versus mean maximum temperature of the warmest month (T_{max}) (black = CG, grey = CHG, open with dotted trend line = CHH). The number of SMVs that were averaged is in parentheses. (e–h) Maps showing the methylation levels per sample (pie charts) in relation to mean maximum temperature (T_{max}; background colour) for the same four examples.

based on several significant associations with climate gradients, in particular for CG-SMVs and the climate variable, T_{max}. Most of these climate-associated CG-SMVs are within or proximal to specific genes, suggesting a potential epigenetic mechanism associated with the expression of those genes (Wang *et al.* 2015). These findings provide support for climate as a force driving the high levels of differentiation of CG-SMVs among three populations of valley oak (Platt *et al.* 2015). In addition, the fact that several of our significant climate-associated SMVs are found in gene bodies is consistent with recent work in *Arabidopsis* showing that gene body CG methylation is functionally important (Takuno & Gaut 2012) and may be involved in local adaptation, either directly or through underlying genetic mechanisms (Dubin *et al.* 2015). Our study joins others in demonstrating that methylation levels differ among naturally occurring plants in different environments (e.g. Herrera & Bazaga 2010; Foust *et al.* 2013), and further offers a genomic context for this methylation, suggesting its connection to local adaptation or tree response to temperature variation.

Patterns of methylation across the genome

Cytosine methylation occurs with different frequencies in CG, CHG and CHH contexts, and these frequencies in valley oak are generally similar to those found in *Arabidopsis* (Cokus *et al.* 2008; Schmitz *et al.* 2013). For example, Schmitz *et al.* (2013) report 23%, 13% and 64% of methylated cytosines were CG, CHG and CHH, respectively, in comparison with our values of 19%, 15% and 66% (Table 1). When restricting to SMVs, defined as sites whose methylation is variable across individuals, we observe that most sites occur in the CG context (46% vs 17% for CHG and 37% for CHH; Table 1), which tend to have a more bimodal methylation distribution than CHG or CHH sites (Fig. 2), similar to *Arabidopsis* (Dubin *et al.* 2015). This high variation in CG-SMVs among samples establishes the potential for its involvement in phenotypic differences among individuals. In contrast, CHH-SMVs are more likely to be unmethylated across all samples, and CHG-SMVs are intermediate (Fig. 2), suggesting these classes are less likely to be involved in phenotypic differences among individuals. These pat-

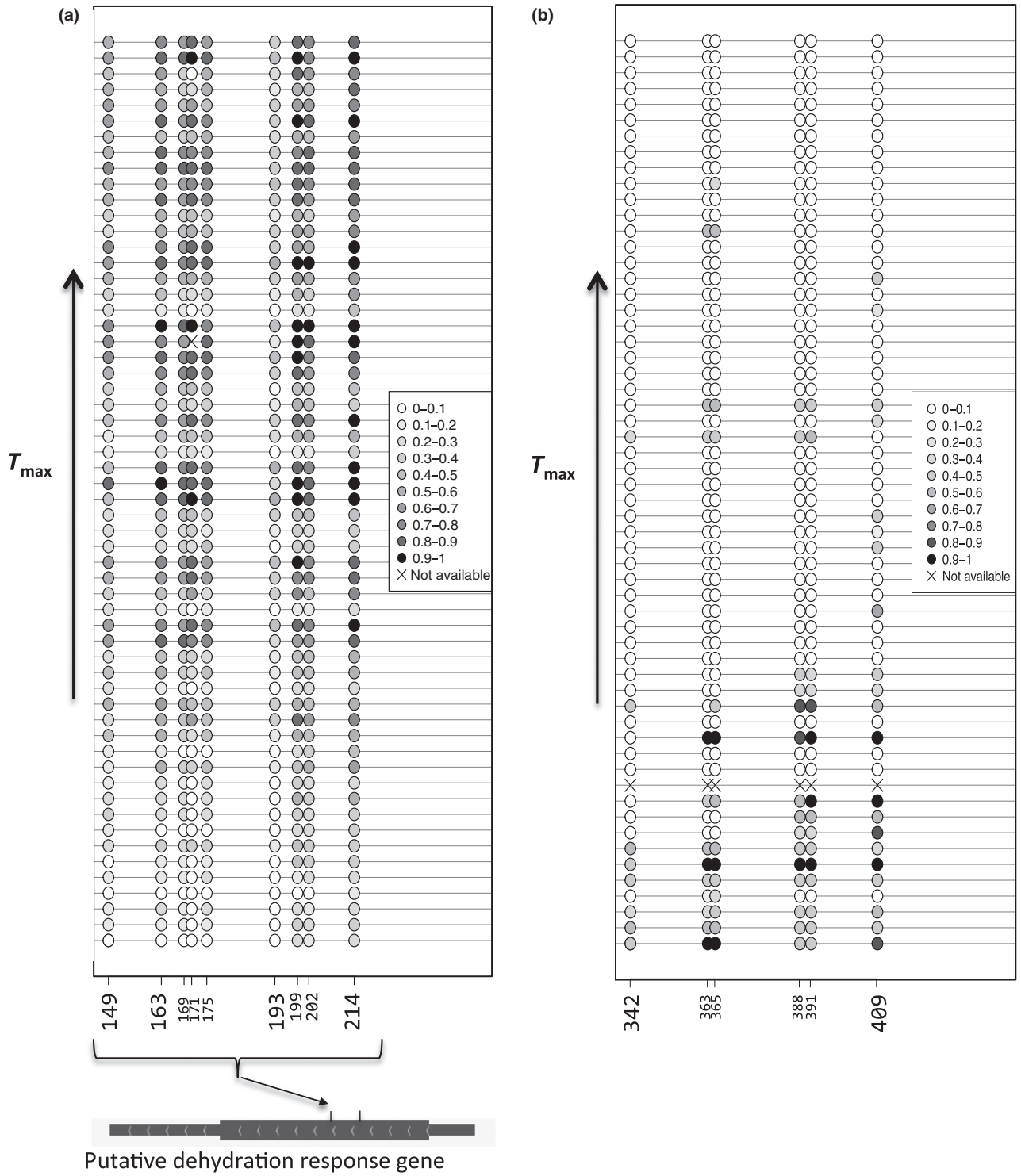

Fig. 5 Methylation levels (grey scale) of (a) nine CG-SMVs from scaffold20751.23308 shown for all samples sorted from high to low T_{max} and shown in their genic context, which is putatively, dehydration-responsive element-binding gene. (b) Similar plot for six CG-SMVs from Locus C2687893 that are intergenic and located 400 bp downstream of a voltage-gated anion channel mitochondrial outer membrane protein. Plots created with METHYLATION PLOTTER (Mallona *et al*. 2014).

Table 4 Number of polymorphisms used for redundancy analyses, per cent variance explained by climate and spatial variables in each association, and the *P*-values for the associations

	Variable sites (no missing data)	% variance explained	*P*
CG-SMV	15 024	14.4	<0.001
CHG-SMV	4382	13.6	<0.001
CHH-SMV	7137	12.8	0.26
SNP	1394	13.3	<0.001

terns are consistent with those observed in an earlier study on three valley oak populations sampled at a smaller spatial scale and using different methods to call methylation (Platt *et al.* 2015).

Climate association with DNA methylation levels

To determine whether climate could be a driver in the previously reported population differentiation at CG sites (Platt *et al.* 2015), we tested the hypothesis that specific SMVs are significantly associated with climate on the landscape. We found that 43 SMVs (from 22 RRBS fragments) are significantly associated with climate variables (Table 3; Fig 3), consistent with their potential role in local adaptation to climate or their plastic response to local climate conditions. Almost all of these variants are CG-SMVs significantly associated with T_{max}, leading to the hypothesis that temperature stress is a selective force or modifying force on CG methylation levels among individuals. The unique geographic patterns of climate-associated SMVs from a sample of different RRBS fragments (Fig. 4) indicate that these forces likely operate independently on different loci.

A notable result is that most significant tests are for associations with maximum temperate (Tables 2 and 3), raising the question of whether temperature is simply a key selective pressure of the local climate environment on CG-SMVs, or whether it is more likely to induce CG-SMV methylation. This finding contrasts with similar analyses in valley oak of SNPs from genes where mean annual precipitation and growing degree-days above 5°C were more important than maximum temperature (Sork *et al.* 2016; Gugger *et al.* in press). A potential cause of strong temperature association is that individual plants acquire CG methylation in response to local temperature conditions during development. For example, temperature can induce epigenetic differentiation during embryogenesis, although the genomic context was not reported (Yakovlev *et al.* 2014). However, research in *Arabidopsis* suggests that CG methylation is more likely to be generated by spontaneous mutation (Becker *et al.* 2011; Hagmann *et al.* 2015), while CHH methylation is more likely to be induced by growing conditions during development and may be mostly reset each generation (Dubin *et al.* 2015). RNA-directed DNA methylation (RdDM) related to transposon silencing can also be temperature sensitive, leading to methylation in all sequence contexts in a given locus (Matzke & Mosher 2014). However, most of the RRBS fragments with significant CG-SMVs do not have adjacent, significant non-CG-SMVs, arguing against RdDM temperature sensitivity as the cause. Finally, there is growing evidence that natural selection can act on DNA methylation, especially CG-SMVs (van der Graaf *et al.* 2015), raising the possibility that temperature associations are due to local adaptation. In this study, it is difficult to determine the reason that maximum temperature is most often associated with CG-SMVs.

Overall, our climate association analyses provide convincing evidence in support of our hypothesis that CG-SMVs, relative to other SMV classes, are more likely to be involved in local adaptation along climate gradients. In addition to the high frequency of significant associations with individual CG-SMVs, multivariate RDA shows that climate and space across sampling localities explain more variance in CG-SMVs (14.4%) than in SNPs (13.3%), CHG-SMVs (13.6%) or CHH-SMVs (12.8%) (Table 4). These differences in explained variance appear modest, but large differences are not necessarily

Table 5 Biplot (Fig. S1, Supporting information) scores indicating the association of individual climate and spatial variables along redundancy analyses axes one (RDA1) and two (RDA2) for CG-SMV, CHG-SMV, CHH-SMV and SNP data sets. The score representing the climate variable with the strongest association with each axis for each redundancy analysis is in bold

	RDA1				RDA2			
	CG	CHG	CHH	SNP	CG	CHG	CHH	SNP
GSDD5	−0.66	−0.55	0.47	0.71	−0.40	−0.66	−0.82	0.65
CWD	−0.45	−0.25	−0.09	0.62	**−0.74**	**−0.79**	**−0.85**	0.24
T_{max}	**−0.86**	**−0.88**	**0.77**	0.22	0.24	−0.34	−0.36	0.57
T_{min}	−0.05	0.07	−0.19	0.60	−0.47	−0.60	−0.74	0.46
Lon	−0.40	−0.10	−0.02	**0.81**	−0.69	−0.56	−0.64	−0.26
Lat	−0.03	−0.31	0.41	−0.58	0.72	0.70	0.63	0.42
Elev	0.14	0.27	−0.22	0.06	−0.08	0.40	0.36	**−0.81**

expected with such large numbers of loci (Lasky *et al.* 2012). Critically, because the entire CG-SMV class has a stronger association with climate than do the SNPs (Tables 4 and 5), which are likely to be neutral on average, CG-SMVs are more likely to be involved in local adaptation to climate than a random locus, whether the mechanism be direct or through linkage to genetic variation (Platt *et al.* 2015) and whether the CG-SMVs arose through spontaneous mutation or environmental induction. CHH-SMVs, on the other hand, appear unlikely to be involved in local adaptation due to their comparatively low climate association. Our results are consistent with research in *Arabidopsis* showing that CG-SMVs are highly associated with latitude and climate on the landscape (Keller *et al.* 2016), whereas CHH-SMVs are associated with response to temperature of greenhouse growing conditions, likely due to their role in transposon silencing (Dubin *et al.* 2015).

Our findings not only point to the role of CG methylation in local plant response to the environment, but this role seems to be different from that of SNPs. In addition to differences in the climate variables shaping SMVs from this study versus SNPs from previous work using mixed model analyses (Sork *et al.* 2016; Gugger *et al.* in press), we also find that SNPs in this study are most strongly associated with spatial variables, whereas SMVs are more strongly associated with climate variables using multivariate analyses (Table 4). These contrasting patterns among SMVs and SNPs establish the potential for adaptive methylation variation to be decoupled from genetic variation and independently targeted by natural selection or other driving forces (van der Graaf *et al.* 2015). Theoretical work suggests that locally adapted phenotypes can arise more rapidly when underlying variation arises both genetically and epigenetically and that variation initially arising by epigenetic mechanisms can eventually lead to traits with a heritable underlying genetic basis (Klironomos *et al.* 2013; Kronholm & Collins 2015).

Candidate genes potentially influenced by DNA methylation

Most of the climate-associated SMVs were found in or within 2.5 kb of a gene (Table 3), suggesting a possible involvement in regulating the expression of those genes and providing a potential mechanism for their involvement in adaptation. Three genes have at least four SMVs associated with T_{max} and are worth highlighting. Most interesting among them, scaffold20751.23308 contains nine CG-SMVs and aligns to a gene identified in both our reference genome and transcriptome to contain an AP2 domain, which is commonly found in transcription factors involved in floral development (Jofuku *et al.* 1994; Okamuro *et al.* 1997). Upon searching the entire NCBI nr database, we found that this locus has high similarity specifically with dehydration-responsive element-binding protein 1A, which mediates transcription under cold, drought and salt stress in *Oryza sativa* (Dubouzet *et al.* 2003). Another interesting locus, C2687893.1977, contains six CG-SMVs and occurs about 400 bp downstream of a gene with high similarity to mitochondrial outer membrane porin 4-like protein in *Solanum* spp. This gene belongs to a family encoding voltage-dependent anion channels in *Arabidopsis* and functions in the regulation of plant growth, development and disease resistance (Tateda *et al.* 2011). Finally, C2243561.395 is unique in having SMVs in all three contexts (CG, CHG and CHH) associated with T_{max}, although it falls on a small (1 kb) genomic contig with no annotation. Locus scaffold 4607.51780 with only a single climate-associated CG-SMV falls within an exon of a gene homologous to a dehydration response protein (dehydRP-like protein) in *Eucalyptus cladocalyx*, which functions in plant–water relations (Bush & Thumma 2013), and is 10 kb downstream from a gene encoding a Staygreen protein (Pfam PF12638; m01oak05463JC), which is implicated in senescence-induced chlorophyll degradation (Park *et al.* 2007). In sum, the climate-associated SMVs can be linked to genes with potential relationships to plant response to the climate environment. Further experimental work on these genes and their expression is necessary to determine how they affect physiological performance.

Is there a genetic basis of climate-associated SMVs?

This research raises the question of whether the climate-associated methylation patterns have a genetic basis elsewhere in the genome, or they independently arise and transmit to offspring. Our data offer an opportunity for preliminary investigation. We used linear mixed models in PyLMM (as above) to test whether variation at any of the 1810 SNPs is significantly associated with the 38 climate-associated CG-SMVs, after adjusting for kinship based on the SNP data. Eight SNPs are highly associated with climate-associated CG-SMVs ($Q < 0.1$). However, the SNPs and CG-SMVs are not near each other along the genome so this analysis does not provide clear evidence for a genetic basis of these CG-SMVs. *Trans*-acting SNPs may underlie a substantial fraction of CG-SMVs in *Arabidopsis,* but the mechanism is not well understood (Dubin *et al.* 2015). Future studies could utilize whole-genome sequences to more thoroughly search for candidate genetic loci associated with methylation sites.

Conclusions

Our study offers evidence in a long-lived tree species that CG-SMVs are involved in plant response to the local environment and possibly the evolution of local adaptation to climate. Empowered by a recently developed draft reference genome sequence, we were able to demonstrate the specific sequence and genic context of climate-associated methylation, offering essential details for how the methylation might be acting. At least three important questions remain unresolved: (i) Are CG-SMVs controlled by underlying genetic variation shaped by natural selection or might they offer an independent, heritable source of variation upon which selection can act? (ii) Are they induced by environmental conditions during the life of the organism, and if so, are they transiently involved in phenotypic plasticity or are they inherited, offering a sort of Lamarckian form of evolution? (iii) Exactly how does this methylation variation influence phenotypes? Whole-genome and whole-epigenome sequencing coupled with experiments examining gene expression and phenotypic response to environmental treatments will provide additional insight into these questions. Nonetheless, our findings indicate that DNA methylation can be a mechanism for plant response to environment leading to local adaptation in natural populations. The implications for response to climate change are profound, especially for long-lived taxa experiencing rapid change, and invite further study of the importance of methylation in other natural ecological and evolutionary contexts.

Acknowledgements

We thank Ana Albarrán, Andy Lentz and Juan Peñaloza for assistance with sampling, Krista Beckley for laboratory work, and Marco Marcello and Liudmilla Rubbi for assistance with the library preparation protocol. We acknowledge the California State Parks and Recreation and the University of California Reserve system for permission to access some of our collection sites and the Broad Stem Cell Research Center for its sequencing facilities. Funding for PFG, SFG and molecular analyses was provided by UCLA as seed money to VLS.

References

Becker C, Weigel D (2012) Epigenetic variation: origin and transgenerational inheritance. *Current Opinion in Plant Biology*, **15**, 562–567.

Becker C, Hagmann J, Muller J *et al.* (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, **480**, 245–249.

Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. *Ecology Letters*, **11**, 106–115.

Bräutigam K, Vining KJ, Lafon-Placette C *et al.* (2013) Epigenetic regulation of adaptive responses of forest tree species to the environment. *Ecology and Evolution*, **3**, 399–415.

Bush D, Thumma B (2013) Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genetics & Genomes*, **9**, 741–752.

Chodavarapu RK, Feng S, Ding B *et al.* (2012) Transcriptome and methylome interactions in rice hybrids. *Proceedings of the National Academy of Sciences*, **109**, 12040–12045.

Cingolani P, Platts A, Wang LL *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, **6**, 80–92.

Cokus SJ, Feng S, Zhang X *et al.* (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219.

Cokus SJ, Gugger PF, Sork VL (2015) Evolutionary insights from *de novo* transcriptome assembly and SNP discovery in California white oaks. *BMC Genomics*, **16**, 552.

Cortijo S, Wardenaar R, Colomé-Tatché M *et al.* (2014) Mapping the epigenetic basis of complex traits. *Science*, **343**, 1145–1148.

Dowen RH, Pelizzola M, Schmitz RJ *et al.* (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences*, **109**, E2183–E2191.

Dubin MJ, Zhang P, Meng D *et al.* (2015) DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife* **4**, e05255.

Dubouzet JG, Sakuma Y, Ito Y *et al.* (2003) OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant Journal*, **33**, 751–763.

Duncan EJ, Gluckman PD, Dearden PK (2014) Epigenetics, plasticity, and evolution: how do we link epigenetic change to phenotype? *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, **322**, 208–220.

Eichten SR, Schmitz RJ, Springer NM (2014) Epigenetics: beyond chromatin modifications and complex genetic regulation. *Plant Physiology*, **165**, 933–947.

Feng S, Rubbi L, Jacobsen SE, Pellegrini M (2011) Determining DNA methylation profiles using sequencing. In: *High-Throughput Next Generation Sequencing: Methods and Applications, Methods in Molecular Biology* (eds Kwon Y. M., Ricke S. C.), pp. 223–238. Humana Press, Totowa, New Jersey.

Flint L, Flint A, Thorne J, Boynton R (2013) Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. *Ecological Processes*, **2**, 25.

Foust CM, Schrey AW, Richards CL (2013) Genetic and epigenetic population structure in *Spartina alterniflora*. *Integrative and Comparative Biology*, **53**, E283–E283.

Foust CM, Preiete V, Schrey AW *et al.* (2016) Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. *Molecular Ecology*, **25**, 1639–1652.

Franks SJ, Hoffmann AA (2012) Genetics of climate change adaptation. *Annual Review of Genetics*, **46**, 185–208.

Gaddis KD, Zukin HL, Dieterich IA, Braker E, Sork VL (2014) Effect of clonal reproduction on genetic structure in *Pentaclethra macroloba* (Fabaceae: Mimosoideae). *Revista de Biología Tropical*, **62**, 443–454.

van der Graaf A, Wardenaar R, Neumann DA *et al.* (2015) Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*, **112**, 6676–6681.

Grivet D, Deguilloux M-F, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology*, **15**, 4085–4093.

Grossniklaus U, Kelly B, Ferguson-Smith AC, Pembrey M, Lindquist S (2013) Transgenerational epigenetic inheritance: how important is it? *Nature Reviews Genetics*, **14**, 228–235.

Gugger PF, Ikegami M, Sork VL (2013) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology*, **22**, 3598–3612.

Gugger PF, Cokus SJ, Sork VL (in press) Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics & Genomes*. doi: 10.1007/s11295-016-0975-1

Guo W, Fiziev P, Yan W *et al.* (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, **14**, 774.

Hagmann J, Becker C, Müller J *et al.* (2015) Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genetics*, **11**, e1004920.

Herrera CM, Bazaga P (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*. *New Phytologist*, **187**, 867–876.

Herrera CM, Bazaga P (2011) Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. *Molecular Ecology*, **20**, 1675–1688.

Herrera CM, Bazaga P (2013) Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. *Botanical Journal of the Linnean Society*, **171**, 441–452.

Holeski LM, Jander G, Agrawal AA (2012) Transgenerational defense induction and epigenetic inheritance in plants. *Trends in Ecology & Evolution*, **27**, 618–626.

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quarterly Review of Biology*, **84**, 131–176.

Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell*, **6**, 1211–1225.

Kang HM, Zaitlen NA, Wade CM *et al.* (2008) Efficient control of population structure in model organism association mapping. *Genetics*, **178**, 1709–1723.

Kang HM, Sul JH, Service SK *et al.* (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, **42**, 348–354.

Keller TE, Lasky JR, Yi SV (2016) The multivariate association between genome-wide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Molecular Ecology*, **25**, 1823–1837.

Klironomos FD, Berg J, Collins S (2013) How epigenetic mutations can affect genetic evolution: model and mechanism. *BioEssays*, **35**, 571–578.

Kronholm I, Collins S (2015) Epigenetic mutations can both help and hinder adaptive evolution. *Molecular Ecology*, **25**, 1846–1858.

Lasky JR, Des Marais DL, McKay JK *et al.* (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, **11**, 204–220.

Li JT, Yang J, Chen DC, Zhang XL, Tang ZS (2007) An optimized mini-preparation method to obtain high-quality genomic DNA from mature leaves of sunflower. *Genetics and Molecular Research*, **6**, 1064–1071.

Liu QA (2013) The impact of climate change on plant epigenomes. *Trends in Genetics*, **29**, 503–505.

Liu Y, Siegmund K, Laird P, Berman B (2012) Bis-SNP: combined DNA methylation and SNP calling for bisulfite-seq data. *Genome Biology*, **13**, R61.

Mallona I, Diez-Villanueva A, Peinado M (2014) Methylation plotter: a web tool for dynamic visualization of DNA methylation data. *Source Code for Biology and Medicine*, **9**, 11.

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics*, **15**, 394–408.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

McLaughlin BC, Zavaleta ES (2012) Predicting species responses to climate change: demography and climate microrefugia in California valley oak (*Quercus lobata*). *Global Change Biology*, **18**, 2301–2312.

Nicotra AB, Segal DL, Hoyle GL *et al.* (2015) Adaptive plasticity and epigenetic variation in response to warming in an Alpine plant. *Ecology and Evolution*, **5**, 634–647.

Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proceedings of the National Academy of Sciences*, **94**, 7076–7081.

Oksanen J, Kindt R, Legendre P *et al.* (2015) vegan: Community Ecology Package.

Orozco LD, Morselli M, Rubbi L *et al.* (2015) Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metabolism*, **21**, 905–917.

Park S-Y, Yu J-W, Park J-S *et al.* (2007) The senescence-induced Staygreen protein regulates chlorophyll degradation. *Plant Cell*, **19**, 1649–1664.

Pederson BS, Eyring K, De S, Yang IV, Schwartz DA (2014) Fast and accurate alignment of long bisulfite-seq reads. *arXiv.org*, arXiv:1401.1129.

Platt A, Gugger PF, Pellegrini M, Sork VL (2016) Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations. *Molecular Ecology*, **24**, 3823–3830.

Plomion C, Aury J-M, Amselem J *et al.* (2016) Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources*, **16**, 254–265.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rehfeldt G (2006) *A spline model of climate for the western United States*. USDA Forest Service, Port Collins, Ohio.

Rico L, Ogaya R, Barbeta A, Peñuelas J (2014) Changes in DNA methylation fingerprint of *Quercus ilex* trees in response to experimental field drought simulating projected climate change. *Plant Biology*, **16**, 419–427.

Schmitz RJ (2014) The Secret Garden-Epigenetic alleles underlie complex traits. *Science*, **343**, 1082–1083.
Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, **334**, 369–373.
Schmitz RJ, Schultz MD, Urich MA *et al.* (2013) Patterns of population epigenomic diversity. *Nature*, **495**, 193–198.
Shen X, De Jonge J, Forsberg SKG *et al.* (2014) Natural *CMT2* variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genetics*, **10**, e1004842.
Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.
Sork VL, Squire KC, Gugger PF *et al.* (2016) Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata* Née (Fagaceae). *American Journal of Botany*, **103**, 33–46.
Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.
Sul JH, Eskin E (2013) Mixed models can correct for population structure for genomic regions under selection. *Nature Reviews Genetics*, **14**, 300–300.
Takuno S, Gaut BS (2012) Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Molecular Biology and Evolution*, **29**, 219–227.
Tateda C, Watanabe K, Kusano T, Takahashi Y (2011) Molecular and genetic characterization of the gene family encoding the voltage-dependent anion channel in *Arabidopsis*. *Journal of Experimental Botany*, **62**, 4773–4785.
Uthup TK, Ravindran M, Bini K, Thakurdas S (2011) Divergent DNA methylation patterns associated with abiotic stress in *Hevea brasiliensis*. *Molecular Plant*, **4**, 996–1013.
Verhoeven KJF, van Gurp TP (2012) Transgenerational effects of stress exposure on offspring phenotypes in apomictic dandelion. *PLoS ONE*, **7**, e38605.
Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytologist*, **185**, 1108–1118.
Walenz B, Florea L (2011) Sim4db and Leaff: utilities for fast batch spliced alignment and sequence indexing. *Bioinformatics*, **27**, 1869–1870.
Wang H, Beyene G, Zhai J *et al.* (2015) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proceedings of the National Academy of Sciences*, **112**, 13729–13734.
Widman N, Feng S, Jacobsen SE, Pellegrini M (2014) Epigenetic differences between shoots and roots in *Arabidopsis* reveals tissue-specific regulation. *Epigenetics*, **9**, 236–242.
Wilschut R, Oplaat C, Snoek B, Kirschner J, Verhoeven KJF (2015) Natural epigenetic variation contributes to heritable flowering divergence in a widespread asexual dandelion lineage. *Molecular Ecology*, **25**, 1759–1768.
Wu T, Watanabe C (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, **9**, 1859–1875.
Yakovlev IA, Asante DKA, Fossdal CG, Junttila O, Johnsen Ø (2011) Differential gene expression related to an epigenetic memory affecting climatic adaptation in Norway spruce. *Plant Science*, **180**, 132–139.
Yakovlev I, Fossdal CG, Skroppa T *et al.* (2012) An adaptive epigenetic memory in conifers with important implications for seed production. *Seed Science Research*, **22**, 63–76.
Yakovlev I, Lee Y, Rotter B *et al.* (2014) Temperature-dependent differential transcriptomes during formation of an epigenetic memory in Norway spruce embryogenesis. *Tree Genetics & Genomes*, **10**, 355–366.
Yoder JB, Stanton-Geddes J, Zhou P *et al.* (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics*, **196**, 1263–1275.
Zhang YY, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytologist*, **197**, 314–322.
Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, **39**, 61–69.

Data accessibility

Illumina sequence reads are available through NCBI BioProject PRJNA306638. The resulting methylation and SNP calls, BAM alignment files, as well as input and output files for redundancy analyses and PyLMM are available through Dryad doi:10.5061/dryad.1b01r. The reference genome sequence Version 0.5 used for this study is available at http://valleyoak.ucla.edu and NCBI BioProject PRJNA308314.

P.F.G. and V.L.S. designed the study, S.F.G. conducted the bioinformatics, P.F.G. and S.F.G. analysed the data with input from V.L.S. and M.P. and P.F.G. and V.L.S. wrote the manuscript with contributions from S.F.G. and M.P.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 QQ-plots for linear mixed models associating CG-SMVs with T_{max} using the methylation data versus SNP data to generate the kinship matrix.

Fig. S2 Scatterplot demonstrating the similarity in results when generating the kinship matrix in PyLMM from the methylation versus genetic data.

Fig. S3 Patterns of average methylation levels and occurrence of methylation sites around transcription start sites.

Fig. S4 Patterns of occurrence of methylation sites and RRBS fragments around transcription start sites.

Fig. S5 Percent of climate-associated SMVs, each class of all SMVs, and each context of all called C sites that are found in

intergenic, exon, intron, upstream <1 kb, or downstream <1 kb regions.

Fig. S6 Triplots showing association of CG-SMV, CHG-SMV, CHH-SMV, and SNP variation with climate and spatial variables based on redundancy analysis (RDA).

Table S1. Sample site coordinates, climate data, and mean methylation levels per sample.

Table S2. Mean correlation of methylation levels within fragments among different classes of methylation contexts